Independent Component analysis of the Cosmic Microwave Background

J-F. Cardoso, J. Delabrouille, G. Patanchon

INDEPENDENT COMPONENT ANALYSIS OF THE COSMIC MICROWAVE BACKGROUND

Jean-François Cardoso

CNRS / ENST - TSI, Paris France

Jacques Delabrouille, Guillaume Patanchon

PCC - Collège de France, Paris, France

ABSTRACT

This paper presents an application of ICA to astronomical imaging. A first section describes the astrophysical context and motivates the use of source separation ideas. A second section describes our approach to the problem: the use of a noisy Gaussian stationary model. This technique uses spectral diversity and takes explicitly into account contamination by additive noise. Preliminary and extremely encouraging results on realistic synthetic signals and on real data will be presented at the conference.

1. COSMIC MICROWAVE BACKGROUND

The cosmic microwave background (CMB), see fig 1 is a 'fossil light' whose observation and characterization has become an essential tool for cosmology, the study of the formation of the Universe. This section reviews a few basic facts about the CMB and explains the relevance of source separation to CMB studies.

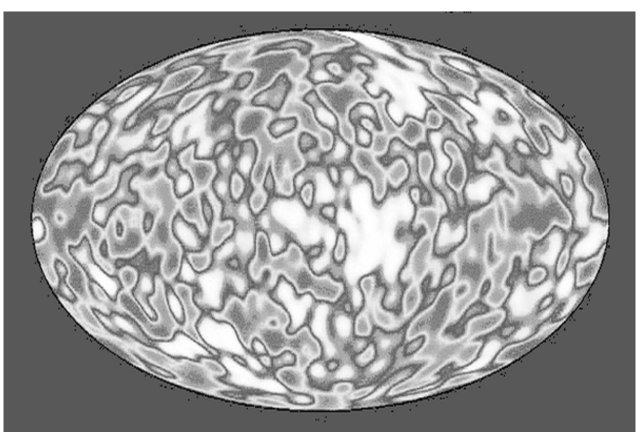

Figure 1: COBE (1989) measurements of the CMB temperature fluctuations over the sky. The amplitude of the fluctuations is of a few micro Kelvins with respect to an average temperature of about 2.726 K.

Big bang The current, widely accepted, model of the formation of the Universe is the 'big bang', a scenario in which the Universe starts very hot and dense and then expands and cools down. In this scenario, the existence of a fossil light has been predicted even before it was observed. The story can be summarized as follows. The time elapsed since the big bang is estimated to be close to 15 billions years. One of the latest milestone events in cosmic life is called the 'atomic recombination' and happened when the Universe was 300,000 years old. Just before atomic recombination, the Universe was extremely homogeneous, its temperature had cooled down to about 3000K, and it was made mostly of photons, electrons and protons (hydrogen nuclei) with a small proportion of other light nuclei. All these particles were strongly interacting and thus were tightly coupled and in thermal equilibrium. However, under the action of a continuing expansion, such a peaceful state of affairs could not go on forever: expansion dilutes energy with a subsequent decrease of temperature. Below a threshold of about 3000K, the thermal agitation is no longer strong enough to prevent 'atomic recombination', that is the formation of atoms from nuclei and electrons. Atoms being electrically neutral, their interaction with light becomes much weaker: all of a sudden (on a cosmic time scale...), photons can freely travel through a transparent Universe: radiation and matter decouple. The photons are set free and for most of them, they are set free forever.

Most of the photons that can be observed *today* have been released at the the time of atomic recombination and have not interacted with matter since then. Thus, the 'cosmic microwave background' is a fossil radiation. It belongs to the microwave domain because photons have cooled down as the Universe itself: their average wavelength, which corresponded to 3000K at recombination time, has been stretched —as the Universe itself— by a factor of about 1000 into the millimetric/centimetric domain. What can be measured today is a radiation with a spectral power density corresponding (to a very high precision) to the radiation of a black body at temperature $T = 2.726$K.

Tiny fluctuations The discovery, by Penzias and Wilson, of a microwave radiation with a black body power spectrum of about 3K, in strong support of the big bang model, is a turning point of modern cosmology. The CMB was found to be extremely homogeneous: the same 3K spectral density was observed to come from all directions in the sky. However, too homogeneous a CMB would be a puzzle in the big bang scenario because the large structures of the Universe (galaxies and galaxy clusters) are thought to have formed via a process of gravitational collapse in which small local over-densities tend to grow by attracting more strongly matter from neighboring regions of lower density. This phenomenon of condensation is in competition with the overall expansion of the Universe which tends to dilute matter. Hence the presence of small inhomogeneities in the original spatial distribution of matter is necessary to understand the latter formation of large scale structures. Because of the coupling between light and matter *before* atomic recombination, small inhomogeneities in matter density should be reflected in the CMB as small inhomogeneities in the spatial distribution of its temperature. Hence, observational cosmology has been feverishly working towards the detection of the *spatial pattern* of CMB temperature in the sky, that is, going beyond its 'first order' feature of being a constant field at $T = 2.726$K. It was not before the COBE space mission in 1989 that instrument sensitivity became high enough to detect tiny fluctuations of a few tens of micro-Kelvins, as displayed in figure 1.

Even though an observational breakthrough, much left to be desired from the COBE experiment because of its limited spatial resolution. As is apparent from the picture, the angular size of the details visible by COBE is of the order of ten degrees. This angular separation corresponds to regions of the Universe which are so large that they cannot have been causally connected in the past! This implies that the low resolution spatial pattern does not carry much information about the *physical* processes which were taking place at recombination time. Thus, in order to extract more information from the fossil light, it is necessary to obtain much finer pictures of CMB anisotropies. Many experiments –ground-based, space-based or balloon-borne– have followed COBE. The ultimate space mission, expected to reach unprecedented levels of accuracy, sensitivity and resolution, is the 'Planck surveyor', whose launch by the European Spatial Agency is planned in 2007. Planck angular resolution is expected to be well under the degree scale. The gain in resolution with respect to COBE is illustrated by the simulation of figure 2. The major difference the two maps of fig. 1 and fig 2 is that the size of the details visible in the COBE map is limited by the resolution of the instruments whereas the typical size of the fluctuation patterns seen on the simulated Planck map is governed by physics: it roughly corresponds to the size of the horizon at recombination time, or about 300,000 light years. There is a conspicuous main peak in the spatial power spectrum of the CMB corresponding to such an angular scale (fig. 3)

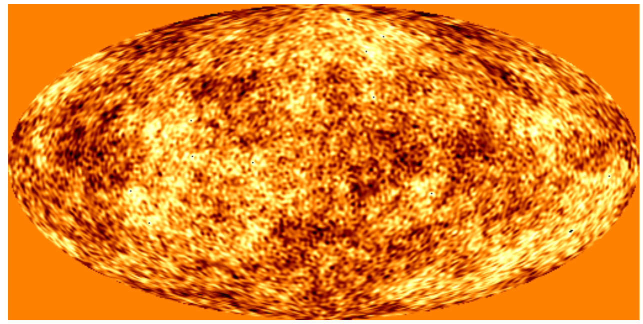

Figure 2: Simulation of CMB anisotropies at expected Planck's resolution and noise (Credit: G.P. Efstathiou)

Harmonic spectrum Gaussian stationary time series are characterized by their power spectrum. For homogeneous isotropic processes on the sphere, the corresponding quantity is called the 'harmonic spectrum'. Denote $\Delta T(\theta, \phi)$ the temperature excess in direction (θ, ϕ). It can be decomposed into *spherical harmonics* as

$$\Delta T(\theta, \phi) = \sum_{l=1}^{\infty} \sum_{m=-l}^{l} a_{lm} Y_{lm}(\theta, \phi), \tag{1}$$

where the doubly indexed set $Y_{lm}(\theta, \phi)$ of spherical harmonic functions is to the sphere what the sines and cosines (of the discrete Fourier transform) are to a one-dimensional interval. A spherical harmonic $Y_{lm}(\theta, \phi)$ accounts for spatial patterns with an angular resolution of $2\pi/l$. For a given multipole l, there are $2l + 1$ spherical harmonics whose coefficients a_{lm} are uncorrelated with variance independent of m when $\Delta T(\theta, \phi)$ is an homogeneous isotropic process on the sphere. The so-called harmonic spectrum, or C_l spectrum, then is $C_l = E\{|a_{lm}|^2\}$ and its empirical estimate is

$$\widehat{C}_l = \frac{1}{2l+1} \sum_{m=-l}^{l} |a_{lm}|^2. \tag{2}$$

Let us stress at this point that the CMB sky stands still (on the human time scale): there is only one Universe and one realization of $\Delta T(\theta, \phi)$ available to us.

Spectral models Cosmologists have constructed models of the CMB formation which predict the shape of the CMB spatial power spectrum as a function of a few 'cosmologic parameters' such as the energy density in the Universe and other quantities of paramount importance to cosmology. An example of the predicted CMB harmonic spectrum is given at figure 3 which is an (appropriately rescaled) plot of the C_l as a function of l. The plot shows in solid line an harmonic spectrum, as

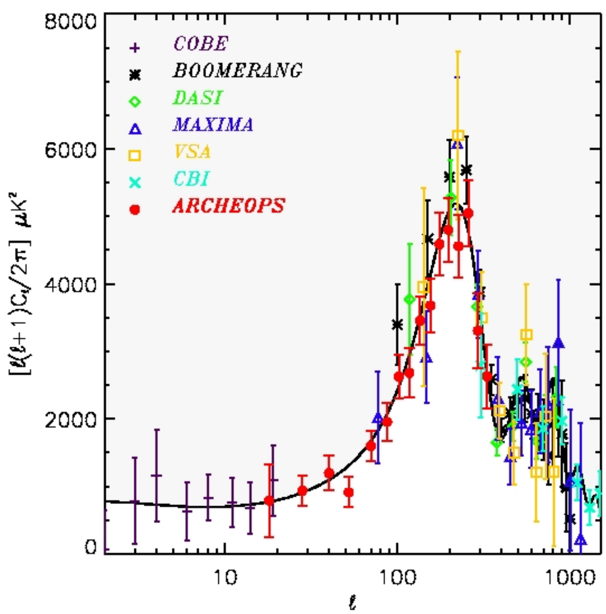

Figure 3: Measurements of the harmonic spectrum C_l by several independent experiments and the best fit by a low dimensional cosmological model.

predicted by the best fit to the data of several CMB experiments. The peaks in the harmonic spectrum are signatures of 'acoustic oscillations', so called because they result from the competition between inertia and elasticity of the medium. This is the reason why the shapes and locations of the peaks directly carry physical information.

The superimposed squares in fig. 3 show how the spectrum is constrained by observations from COBE and Boomerang, a balloon-borne experiment which flew over the North pole in 1998. Note that the existence of a main peak is confirmed by the experiments.

Astronomical components The CMB is not expected to be the only emission visible in the centimeter range. Contributions of other sources are called *foregrounds* and should include Galactic dust emission, emission from very remote (and hence quasi point-like) galaxy clusters, and several others, including instrumental effects. Figure 4 illustrates the situation.

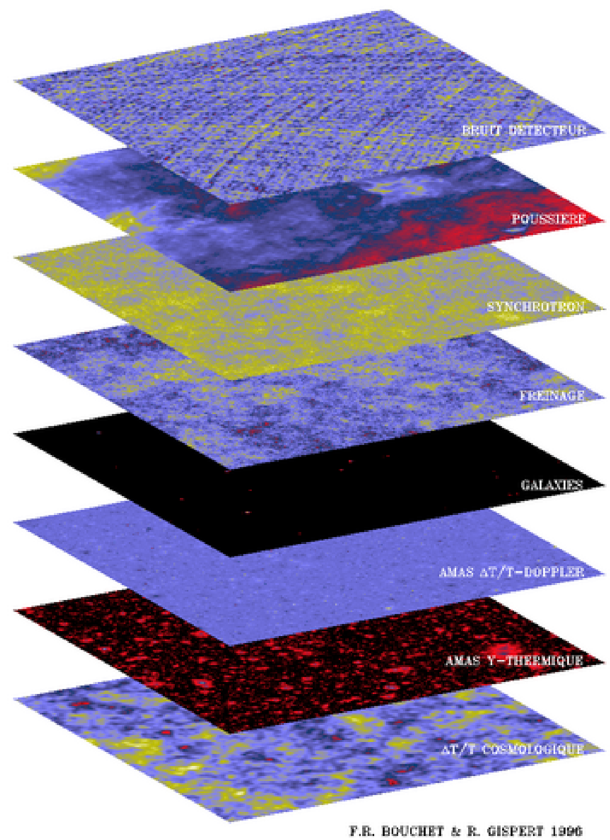

Figure 4: Several emissions contributes to the sky map in addition to the CMB. Credit: Bouchet & Gispert

Separating the cosmic background from the foregrounds is possible ICA style, because the instruments perform *multi-spectral* imaging, meaning that sky maps are obtained in several frequency bands (or channels). The good news for ICA people is that the map built from the i-th channel is expected to be well modeled (after some heavy pre-processing) as

$$\Delta T_i(\theta, \phi) = \sum_j A_{ij} S_j(\theta, \phi) + V_i(\theta, \phi)$$

where $S_j(\theta, \phi)$ is the spatial pattern for the j-th component and $V_i(\theta, \phi)$ is an additive detector noise. In other words, cosmologists expect to observe a *noisy instantaneous (non convolutive) mixture* of components. Further, these components should be statistically independent due to their physically distinct origins.

Blind separation methods This paper is concerned with *blind* component separation. The motivation for a blind approach is obvious: even though some coefficients of the mixture may be known in advance with good accuracy (in particular those related to the CMB), some other components are less well known or predictable. It is thus very tempting to run blind algorithms which do not require *a priori* information about the mixture coefficients.

Several attempts at blind component separation for CMB imaging have already been reported. The first proposal, due to Baccigalupi *et al.* use a non Gaussian noise-free i.i.d. model for the components[3], hence following the 'standard' path to source separation. One problem with this approach is that the most important component, namely the CMB itself, closely follows a Gaussian distribution. It is well known that, in i.i.d. models, it is possible to accommodate at most one Gaussian component. It seems hazardous, however, to use a non Gaussian model when the main component itself has a Gaussian distribution.

Another reason why the i.i.d. modeling (which is implicit in 'standard' ICA) probably is not appropriate to our application: most of the components are very much dominated by the low-frequency part of their spectra. Thus sample averages taken through the data set tend not to converge very quickly to their expected values.

Section 2 describes a generic method for the blind separation of noisy mixtures of stationary processes, based on a spectral approximation to the likelihood which seems to work very well on CMB data (see sec. 3).

2. BLIND SEPARATION OF NOISY MIXTURES VIA SPECTRAL MATCHING

We describe a 'spectral matching' method for the blind separation of noisy mixtures of stationary sources. For ease of exposition, we explain the method as applied to times series rather than to images. Extension to images is straightforward (see sec. 3).

The noisy stationary Gaussian model We model an $m \times 1$-dimensional process $y(t) = [y_1(t); \ldots ; y_m(t)]$ as

$$y(t) = As(t) + v(t) \tag{3}$$

where A is an $m \times n$ matrix with linearly independent columns. The n-dimensional source process $s(t)$ (the components) and the m-dimensional noise process $v(t)$ are modeled as real valued, mutually independent and stationary with spectra $R_s(\nu)$ and $R_v(\nu)$ respectively. The observed process then has spectrum

$$R_y(\nu) = AR_s(\nu)A^\dagger + R_v(\nu). \tag{4}$$

Matrices $R_y(\nu)$, $R_s(\nu)$ and $R_v(\nu)$ are sometimes called 'spectral covariance matrices'. Independence between components implies that $R_s(\nu)$ is a diagonal matrix:

$$[R_s(\nu)]_{ij} = \delta_{ij} P_i(\nu) \quad 1 \le i, j \le n$$

where $P_i(\nu)$ is the power spectrum of the ith source at frequency ν and δ_{ij} is the Kronecker symbol. For simplicity, we also assume that the observation noise is uncorrelated across sensors:

$$[R_v(\nu)]_{ij} = \delta_{ij} V_i(\nu) \quad 1 \le i, j \le m \tag{5}$$

Spectral statistics. A key feature of our method is that it uses low dimensional statistics obtained as averages over *spectral domains* in Fourier space. In the 1D case, spectral domains simply are frequency bands. Let $(-\frac{1}{2}, \frac{1}{2}) = \cup_{q=1}^{Q} \mathcal{D}_q$ be a partition of the frequency interval $(-\frac{1}{2}, \frac{1}{2})$ into Q domains (bands). We require them to be symmetric: $f \in \mathcal{D}_q \Rightarrow -f \in \mathcal{D}_q$.

For any function $g(\nu)$ of frequency, denote $\langle g \rangle_q$ its average over the q-th spectral domain when sampled at multiples of $1/T$:

$$\langle g \rangle_q = \frac{1}{w_q} \sum_{\frac{p}{T} \in \mathcal{D}_q} g\left(\frac{p}{T}\right) \quad q = 1, \ldots, Q \tag{6}$$

where w_q is the number of points in domain $\mathcal{D}_q$.

Denoting $Y(\nu)$ the DFT of T samples:

$$Y(\nu) = \frac{1}{\sqrt{T}} \sum_{t=0}^{T-1} y(t) \exp(-2i\pi\nu t), \tag{7}$$

the *periodogram* is $\widehat{R}_y(\nu) = Y(\nu)Y(\nu)^\dagger$ and its averaged version is

$$\langle \widehat{R}_y \rangle_q = \langle Y(\nu)Y(\nu)^\dagger \rangle_q. \tag{8}$$

Note that $Y(-\nu) = Y(\nu)^*$ for real data so that $\langle \widehat{R}_y \rangle_q$ actually is a real valued matrix if $\mathcal{D}_q$ is a symmetric domain.

This sample spectral covariance matrix will be our estimate for the corresponding averaged quantity

$$\langle R_y \rangle_q = A \langle R_s \rangle_q A^\dagger + \langle R_v \rangle_q \tag{9}$$

where equality (9) results from averaging (4). The structure of the model is not affected by spectral averaging in the sense that $R_s(\nu)$ and $R_v(\nu)$ remain diagonal matrices after averaging:

$$\langle R_s \rangle_q = \mathrm{diag}\,[\langle P_1 \rangle_q, \ldots, \langle P_n \rangle_q] \tag{10}$$

$$\langle R_v \rangle_q = \mathrm{diag}\,[\langle V_1 \rangle_q, \ldots, \langle V_m \rangle_q] \tag{11}$$

Blind identification via spectral matching Our proposal for blind identification simply is to match the sample spectral covariance matrices $\langle \widehat{R}_y \rangle_q$, which depend on the data, to their theoretical values $\langle R_y \rangle_q$, which depend on the parameters to be estimated.

It is up to the user to decide what the 'unknown parameters' are. For instance, one may assume that the noise is temporally white so that $\langle V_i \rangle_q = \sigma_i^2$ for all $q = 1, Q$. Similarly, if the spectrum of the i-th component is known in advance, there is no need to include $\langle P_i \rangle_q$ in the adjustable parameter set. Whatever constraints are imposed on A, on $\langle P_i \rangle_q$ or on $\langle V_i \rangle_q$, we denote by θ the set of parameters needed to represent the unknown quantities in $(A,\, \langle P_i \rangle_q,\, \langle V_i \rangle_q)$.

The unknown quantities θ are estimated by minimizing the following mismatch measure

$$\phi(\theta) = \sum_{q=1}^{Q} w_q \; D\left(\langle \widehat{R}_y \rangle_q, \langle R_y \rangle_q \right) \qquad (12)$$

between the sample statistics and their expected values, where $D(\cdot, \cdot)$ is the measure of divergence between two $m \times m$ positive matrices defined as

$$D(R_1, R_2) = \mathrm{tr}\left(R_1 R_2^{-1} \right) - \log\det(R_1 R_2^{-1}) - m \quad (13)$$

(which is nothing but the Kullback divergence between two m-dimensional zero-mean Gaussian distributions with positive covariance matrices R_1 and R_2.)

Thus, the data are summarized by the matrix set $\langle \widehat{R}_y \rangle_1, \ldots, \langle \widehat{R}_y \rangle_Q$ and the dependence of $\phi(\theta)$ on θ is due to θ determining (by definition) A, $\langle P_i \rangle_q$, and $\langle V_i \rangle_q$ which, in turn, determine $\langle R_y \rangle_q$ via eqs (9-10-11).

The reason for using the mismatch measure (12) is its connection to maximum likelihood principle. Minimizing (12) is equivalent to maximizing the likelihood of the observations in a Gaussian stationary model where the DFT coefficients are approximated as being normally and independently distributed. This is known as the 'Whittle approximation' and and has been used for noise-free ICA by Pham [5]. In the noise-free case, the objective (12) boils down to a joint diagonalization criterion which can be optimized very efficiently. Algorithms for the noisy case are briefly discussed next.

Algorithmics We only hint at the algorithm issue of minimizing the mismatch (12); more details can be found in [4].

Because of its connection to likelihood, criterion (12) can be optimized using the EM algorithm. In contrast to the case of noisy mixtures of *non Gaussian* sources, the Gaussian criterion (12) is easily minimized with EM. Two key points contribute to the simplicity of a spectral domain EM. First, the conditional expectations which appear in EM are *linear* functions of the data in the Gaussian model. Second, this linearity is preserved through domain averaging, meaning that EM only needs to operate on the sample covariance matrices $\langle \widehat{R}_y \rangle_q$. This set of matrices form a sufficient statistic in our model; it is all that is needed to run the EM algorithm. It is also of much lower size than the data themselves (for reasonable choices of the spectral domains).

The EM algorithm provides us with a simple algorithm for the minimization of (12) but it may be slow in the *finishing* phase. This is true in particular in the case of CMB data for which the components are well below the noise level in high frequency bands, that is, $A_{ij}^2 P_j(q) \ll V_i(q)$ for some channel i and some component j for the largest q's. In this case, $P_j(q)$ contributes very little to $\langle R_y \rangle_q$, so that its variations have a very weak impact on the likelihood. Thus, we have found necessary in our application, to complete the minimization of (12) with a quasi-Newton method. The EM is used to provide us with a very good starting point so that the quasi-Newton descent (we use the BFGS method) can be initialized in a favorable situation.

3. APPLICATION TO CMB DATA

In the CMB application, one must process mixtures of images. The spectral matching method applies just as well for 'small' images, with the only modification that the spectral domains $\mathcal{D}_q$ now are areas in the bi-dimensional Fourier plane rather than just spectral bands. The most natural choice for CMB data analysis is to use 'spectral rings' because we expect isotropic spectra but one may elect to use other shapes like *fractions* of spectral rings. This could be a sensible choice in the presence of 'stripes': these are directional artifacts (due the sky-scanning strategy) and, if they are believed to significantly pollute the data set, they can be excluded from the averaging whenever they are well localized in the Fourier plane.

When processing 'large' areas of the sky, one may no longer ignore curvature effects: the 2D Fourier transform must be replaced by an harmonic transform as in (1) and the spectral domains should logically include all (l, m) pairs for a given l. The only effect, from the algorithmic point of view, of using an harmonic transform is that it takes longer to compute than a 2D Fourier transform. Once the sample covariance matrices $\langle \widehat{R}_y \rangle_q$ are computed, the method is identical.

Room is lacking to include figures illustrating the performance of the method. The interested reader is referred to [4] for a performance report on realistic synthetic data sets. Results on real data coming from the balloon-borne Archeops mission (2002) will be presented at the conference.

Multi-detector multi-component spectral estimation. In ICA, the parameters governing the source distributions are usually considered as 'nuisance parameters' as opposed to the 'parameter of interest' which is the mixing matrix. Indeed, in the noise-free case, an estimate of the mixing matrix is all that is needed to invert the mixture and recover the sources. The situation is different in the noisy case because the sources are best estimated by taking the noise into account. In our experiments, we use Wiener filtering in the frequency domain to estimate the sky maps of the sources. Implementing the Wiener filter requires to know the spectra of the sources and of the noise; these quantities are precisely those which are estimated (as 'nuisance parameters') in the minimization of (12).

It should be stressed that our spectral matching method is of interest even in the 'non-blind' case, that is, when the mixing matrix A is *known*, because the minimization of the spectral mismatch (12) with respect to the source spectra $P_i(q)$ (and also, possibly with respect to the noise $V_(q)$) for a fixed value of A yields a non parametric estimate of the source spectra. These are maximum likelihood estimates (in the Whittle approximation) which take into account the *joint information* provided by all the channels and these estimates are obtained without even separating the sources. This is to be compared with a more naive approach to the estimation of the source spectral density in which one would first try to separate the sources and then perform regular univariate spectral estimation. In this case, it is not clear how to take into account the effects of the Wiener filtering and how to remove the effects of the additive noise. In contrast, the spectral matching method directly yields ML spectral estimates for all components.

4. CONCLUSION

Blind source separation appears as a promising tool for CMB studies. We have proposed a method which exploits spectral diversity and boils down to adjusting smoothed versions of the spectral covariance matrices (4) to their empirical estimates.

The matching criterion is equivalent to the likelihood in the Whittle approximation, providing us with a principled way of exploiting the spectral structure of the process and leading to the selection of (12) as an objective function. This objective function uses a summary of the data in the form of a set of average spectral covariance matrices, offering large computational savings, especially when dealing with images.

Implementing the EM algorithm on spectral domains offers a simple method to minimize the objective function but it may be necessary to complement it with a specialized optimization method. Our approach is able to estimate all the parameters: mixing matrix, average source spectra, noise level in each sensor. These parameters can then be used to reconstruct the sources (if needed) via Wiener filtering.

We conclude by noting that, in the presence of non Gaussian data, the Whittle approximation remains a valid tool (the minimization of (12) still yields consistent estimates) but it does *not* capture all the probability structure, even in large samples. Thus, more work remains to be done, in particular to exploit the high non Gaussianity of some components (galaxy clusters appears as quasi point-like structures at the current resolutions).

Resources For the interested reader, many wonderful sites are available on the Internet to get started with the big bang, the Planck mission and everything cosmologic. Among those, I have particularly appreciated Ned Wright's cosmology tutorial [2] (with its section 'News of the Universe') and the extensive material maintained by Wayne Hu on his home page [1] at the University of Chicago.

5. REFERENCES

[1] The home page of Wayne Hu at the University of Chicago. http://background.uchicago.edu/∼whu/.

[2] Ned wright's cosmology tutorial. http://www.astro.ucla.edu/∼ wright/cosmolog.htm.

[3] C. Baccigalupi, L. Bedini, et al. Neural networks and separation of cosmic microwave background and astrophysical signals in sky maps. *Monthly Notices of the Royal Astronomical Society*, 318:769–780, November 2000.

[4] J. Delabrouille, J.-F. Cardoso, and G. Patanchon. Multi–detector multi–component spectral matching and applications for CMB data analysis. Submitted to *Monthly Notices of the Royal Astronomical Society*, 2003. Available as http://arXiv.org/abs/astro-ph/0211504

[5] Dinh-Tuan Pham. Blind separation of instantaneous mixture of sources via the gaussian mutual information criterion. In *Proc. EUSIPCO*, pages 621–624, 2000.